

Two Simple Tests for Heteroscedastic Two-Way ANOVA

Xuefeng Liu¹, Jia Guo², Bu Zhou³ and Jin-Ting Zhang^{*4}

¹IBM, Xi-An, China

^{2,3,*4}Dept of Statistics & Applied Probability, National University of Singapore, Singapore 117546

¹liuxuefeng@gmail.com; ²a0095904@u.edu.sg; ³bu.zhou@u.edu.sg; ^{*4}stazjt@nus.edu.sg

Abstract

Heteroscedastic two-way ANOVA models are frequently encountered in experimental sciences, e.g., biology, psychology and physics among others. In the literature, classical F-tests are often blindly employed although they are often biased even for moderate heteroscedasticity. To overcome this problem, several approximate tests, e.g., generalized P-values, bootstrap, and permutation-based tests among others, have been proposed in the literature. These tests, however, are computationally intensive or do not work well in terms of size controlling and power. In this paper, we study tests of linear hypotheses in heteroscedastic two-way ANOVA via proposing a modified Bartlett (MB) test and a parametric bootstrap (PB) test. The MB test is easy to implement via using the usual χ^2 -distribution and it is shown to be invariant under affine transformations, different choices of the contrast matrix used to define the same hypothesis and different labelling schemes of the cell means. The PB test, on the other hand, is easy to implement but it is time-consuming. Simulations and real data applications show that the MB test and the PB test are comparable and they both outperform the classical F-test in terms of size controlling and power for various sample sizes and parameter configurations.

Keywords: Heteroscedastic Two-way ANOVA; Modified Bartlett Correction; Tests of Linear Hypotheses; Wald-type Statistic

1 Introduction

A two-way analysis of variance (ANOVA) aims to compare the effects of several levels of two factors in a factorial experiment with two-way layout. It is widely used in experimental sciences, e.g., biology, psychology, physics, among others. There is a vast literature concerned with two-way ANOVA using classical F-tests. A good survey is given by Fujikoshi (1993). The classical F-test works well when the cell variances are homogeneous. However, when this homogeneity assumption is not satisfied, the classical F-test is either too conservative (the empirical size is much smaller than the nominal one) or too liberal (the empirical size is much larger than the nominal one) as demonstrated in Table 2.

In real data analysis, the homogeneity assumption for cell variances is often difficult to check and is often not satisfied. If this is the case, two-way ANOVA is usually referred to as heteroscedastic two-way ANOVA. An application of the classical F-test to the main and interaction effect tests of heteroscedastic two-way ANOVA may lead to misleading conclusions (Wilcox 1989, Krutchkoff 1989, Ananda and Weerahandi 1997). To solve this problem, several approximation tests have been proposed. Wilcox (1989) proposed two such tests. The first one generalized James (1954)'s second order test. Since James's second order test has an extremely complicated form, so is Wilcox's version. Wilcox's (1989) second test is not invariant with respect to permutation and it has a low power according to the simulation results presented in Wilcox (1989). Krutchkoff (1989) proposed a simulation-based approximate test. Ananda and Weerahandi (1997) proposed another simulation-based test using the concept of generalized P-values, resulting in a so-called generalized F-test. Recently, some more simulation-based approaches have been proposed for ANOVA models. Bao and Ananda (2001) and Mu et al. (2008) gave further studies on the generalized P-values method. Xu et al. (2013) proposed a parametric bootstrap test and compared it with the generalized F-test of Ananda and Weerahandi (1997). More recently, Konietschke et al. (2015) proposed a bootstrap method while Pauly et al. (2015) studied a permutation test for general ANOVA problems.

Note that the above simulation-based tests are generally computationally intensive or do not work well in terms of size controlling and power. Therefore, it is still of interest to propose and study some simple and accurate tests for the main and interaction effect tests of heteroscedastic ANOVA or MANOVA. Much work has been done in this direction. Examples include Zhang (2012a,b) who studied an approximate degrees of freedom F-test for heteroscedastic two-way ANOVA and one-way MANOVA respectively, and Harrar and Bathke (2010) who proposed modified Wilks likelihood ratio, Lawley-Hotelling trace and Bartlett-Nanda-Pillai tests for heteroscedastic two-way MANOVA among others. Zhang (2011) noted that Harrar and Bathke's modified MANOVA tests are not affine invariant. To address this problem, he proposed an approximate Hotelling T^2 test using Wishart-approximation. Harrar and Bathke (2010)'s modified MANOVA tests were further improved by Zhang and Xiao (2012).

In this paper, alternatively, we propose and study two new tests, a modified Bartlett (MB) test and a parametric bootstrap (PB) test for the general linear hypothesis testing (GLHT) problem in heteroscedastic two-way ANOVA which includes the main,

interaction effect, post hoc, and contrast tests among others as special cases. The MB test is obtained via an application of the modified Bartlett correction of Fujikoshi (2000) to a Wald-type statistic constructed for the GLHT problem. The MB test is easy to implement via the usual χ^2 -distribution and it is invariant under affine transformations, different choices of the contrast matrix used to define the same hypothesis and different labelling schemes of the cell means. The PB test, on the other hand, is easy to implement but it is time-consuming. Simulation studies and real data applications presented in this paper demonstrate that (1) the PB and MB tests are comparable in terms of size controlling and power but the MB test requires much less computational work than the PB test; and (2) the classical F test is very liberal or conservative for heteroscedastic two-way ANOVA in many cases of the simulation studies presented in this paper.

The rest of the paper is organized as follows. The methodologies for the PB and MB tests are presented in Section 2. Simulation results are presented in Section 3. A real data example is presented in Section 4. Finally, some technical proofs of the main results are given in the Appendix.

2 Methodologies

2.1 Tests of Main and Interaction Effects in Heteroscedastic Two-way ANOVA

Consider a two-way experiment with two factors A and B having a and b levels respectively with a total of ab factorial combinations or cells. Suppose at the (i, j) -th cell, we have a random sample: $y_{ijk}, k = 1, 2, \dots, n_{ij}$ satisfying the following model:

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim N(0, \sigma_{ij}^2), \quad k = 1, \dots, n_{ij}, \quad (1)$$

where μ_{ij} and σ_{ij}^2 are the associated cell mean and variance. All these ab samples are assumed to be independent with each other but there is no knowledge about if the cell variances $\sigma_{ij}^2, i = 1, \dots, a; j = 1, \dots, b$ are equal to each others.

For a two-way ANOVA model, the cell means μ_{ij} are usually decomposed into the form $\mu_{ij} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij}, i = 1, 2, \dots, a; j = 1, 2, \dots, b$, where μ_0 is the grand mean, α_i and β_j are the i -th and j -th main effects of factors A and B respectively, and γ_{ij} is the (i, j) -th interaction effect between factors A and B so that (1) can be further written as the following well-known two-way ANOVA model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim N(0, \sigma_{ij}^2), \quad k = 1, 2, \dots, n_{ij}; \quad i = 1, 2, \dots, a; j = 1, 2, \dots, b. \quad (2)$$

For this model, interested are the following three null hypotheses:

$$\begin{aligned} H_{0A} &: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0, \\ H_{0B} &: \beta_1 = \beta_2 = \dots = \beta_b = 0, \\ H_{0AB} &: \gamma_{11} = \dots = \gamma_{1b} = \dots = \gamma_{a1} = \dots = \gamma_{ab} = 0. \end{aligned} \quad (3)$$

The first two null hypotheses aim to test if the main effects of the two factors are significant while the last one aims to test if the interaction effect between the two factors is significant. The model (2) is not identifiable since the parameters μ_0, α_i, β_j and γ_{ij} are not uniquely defined unless further constraints are imposed. Given a sequence of positive weights $w_{ij}, i = 1, 2, \dots, a; j = 1, 2, \dots, b$, we impose the following constraints:

$$\sum_{i=1}^a w_{i\cdot} \alpha_i = 0, \quad \sum_{j=1}^b w_{\cdot j} \beta_j = 0, \quad (4)$$

$$\sum_{i=1}^a w_{ij} \gamma_{ij} = 0, \quad j = 1, 2, \dots, b-1, \quad (5)$$

$$\sum_{j=1}^b w_{ij} \gamma_{ij} = 0, \quad i = 1, 2, \dots, a-1, \quad (6)$$

$$\sum_{i=1}^a \sum_{j=1}^b w_{ij} \gamma_{ij} = 0, \quad (7)$$

where $w_{i\cdot} = \sum_{j=1}^b w_{ij}$ and $w_{\cdot j} = \sum_{i=1}^a w_{ij}$. Notice that we here use only $a + b + 1$ constraints that imply the $a + b + 2$ constraints suggested by Fujikoshi (1993) and adopted by Ananda and Weerahandi (1997). This is because (5) jointly with (7) implies that (6) holds for $i = a$ while (6) jointly with (7) implies that (5) holds for $j = b$. Set $\alpha = [\alpha_1, \dots, \alpha_a]^T, \beta = [\beta_1, \dots, \beta_b]^T$, and $\gamma = [\gamma_{11}, \dots, \gamma_{1b}, \dots, \gamma_{a1}, \dots, \gamma_{ab}]^T$. Then under (4)~(7), simple algebra shows that the three null hypotheses (3) can be equivalently written as

$$\begin{aligned} H_{0A} &: \mathbf{H}_a \alpha = 0, \quad \text{where } \mathbf{H}_a = \begin{pmatrix} \mathbf{I}_{a-1} & -\mathbf{1}_{a-1} \end{pmatrix}, \\ H_{0B} &: \mathbf{H}_b \beta = 0, \quad \text{where } \mathbf{H}_b = \begin{pmatrix} \mathbf{I}_{b-1} & -\mathbf{1}_{b-1} \end{pmatrix}, \\ H_{0AB} &: \mathbf{H}_{ab} \gamma = 0, \quad \text{where } \mathbf{H}_{ab} = \mathbf{H}_a \otimes \mathbf{H}_b, \end{aligned} \quad (8)$$

where and throughout, $\mathbf{I}_r$ and $\mathbf{1}_r$ denote the identity matrix of size r and the r -dimensional vector of ones respectively, and $\otimes$ denotes the Kronecker product operation. The matrices $\mathbf{H}_a$, $\mathbf{H}_b$ and $\mathbf{H}_{ab}$ are full rank contrast matrices whose rows sum up to 0. The contrast matrices $\mathbf{H}_a$, $\mathbf{H}_b$ and $\mathbf{H}_{ab}$ are not unique. For example, $\tilde{\mathbf{H}}_a = \begin{pmatrix} -1_{a-1} & \mathbf{I}_{a-1} \end{pmatrix}$ is also a contrast matrix associated with the null hypothesis H_{0A} . However, we later will show that the MB and PB tests, which will be proposed in this paper, will not be affected by different choices of the contrast matrices.

When the weights can be written as $w_{ij} = u_i v_j$, $i = 1, 2, \dots, a; j = 1, 2, \dots, b$ such that $u_i > 0$, $\sum_{i=1}^a u_i = 1$, and $v_j > 0$, $\sum_{j=1}^b v_j = 1$, we can easily identify the parameters μ_0, α_i, β_j , and γ_{ij} as $\mu_0 = \sum_{i=1}^a \sum_{j=1}^b u_i v_j \mu_{ij}$, $\alpha_i = \sum_{j=1}^b v_j \mu_{ij} - \mu_0$, $\beta_j = \sum_{i=1}^a u_i \mu_{ij} - \mu_0$, and $\gamma_{ij} = \mu_{ij} - \alpha_i - \beta_j - \mu_0$. Let $\mathbf{u} = [u_1, \dots, u_a]^T$, $\mathbf{v} = [v_1, \dots, v_b]^T$ and $\boldsymbol{\mu} = [\mu_{11}, \dots, \mu_{1b}, \dots, \mu_{a1}, \dots, \mu_{ab}]^T$. Denote the s -dimensional unit vector whose r -th component is 1 and others are 0 as $\mathbf{e}_{r,s}$. Then we further have $\mu_0 = [\mathbf{u}^T \otimes \mathbf{v}^T] \boldsymbol{\mu}$, $\alpha_i = [(\mathbf{e}_{i,a} - \mathbf{u})^T \otimes \mathbf{v}^T] \boldsymbol{\mu}$, $\beta_j = [\mathbf{u}^T \otimes (\mathbf{e}_{j,b} - \mathbf{v})^T] \boldsymbol{\mu}$, and $\gamma_{ij} = [(\mathbf{e}_{i,a} - \mathbf{u})^T \otimes (\mathbf{e}_{j,b} - \mathbf{v})^T] \boldsymbol{\mu}$, $i = 1, 2, \dots, a; j = 1, 2, \dots, b$. In matrix notation, we can further write

$$\boldsymbol{\alpha} = \mathbf{A}_a \boldsymbol{\mu}, \quad \boldsymbol{\beta} = \mathbf{A}_b \boldsymbol{\mu}, \quad \text{and} \quad \boldsymbol{\gamma} = \mathbf{A}_{ab} \boldsymbol{\mu},$$

where $\mathbf{A}_a = (\mathbf{I}_a - \mathbf{1}_a \mathbf{u}^T) \otimes \mathbf{v}^T$, $\mathbf{A}_b = \mathbf{u}^T \otimes (\mathbf{I}_b - \mathbf{1}_b \mathbf{v}^T)$, and $\mathbf{A}_{ab} = (\mathbf{I}_a - \mathbf{1}_a \mathbf{u}^T) \otimes (\mathbf{I}_b - \mathbf{1}_b \mathbf{v}^T)$.

There are a few methods which can be used to specify the weights w_{ij} , $i = 1, 2, \dots, a; j = 1, 2, \dots, b$; see for example, Fujikoshi (1993). In this paper, we shall use the following two simple methods: the equal-weight method and the size-adapted-weight method. Both the methods specify the weights as $w_{ij} = u_i v_j$, $i = 1, 2, \dots, a; j = 1, 2, \dots, b$ with the equal-weight method specifying $\mathbf{u}$ and $\mathbf{v}$ as $u_i = 1/a, v_j = 1/b$, $i = 1, 2, \dots, a; j = 1, 2, \dots, b$ while the size-adapted-weight method specifying $\mathbf{u}$ and $\mathbf{v}$ as $u_i = \sum_{j=1}^b n_{ij}/N$, $i = 1, 2, \dots, a$ and $v_j = \sum_{i=1}^a n_{ij}/N$, $j = 1, 2, \dots, b$ where $N = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$ denotes the total sample size. Under the equal-weight method, the matrices $\mathbf{A}_a$, $\mathbf{A}_b$ and $\mathbf{A}_{ab}$ have the following simple forms:

$$\mathbf{A}_a = (\mathbf{I}_a - \frac{\mathbf{J}_a}{a}) \otimes \frac{\mathbf{1}_b}{b}, \quad \mathbf{A}_b = \frac{\mathbf{1}_a}{a} \otimes (\mathbf{I}_b - \frac{\mathbf{J}_b}{b}), \quad \mathbf{A}_{ab} = (\mathbf{I}_a - \frac{\mathbf{J}_a}{a}) \otimes (\mathbf{I}_b - \frac{\mathbf{J}_b}{b}),$$

where $\mathbf{J}_m = \mathbf{1}_m \mathbf{1}_m^T$ is a $m \times m$ matrix of ones. When the two-way ANOVA design is balanced, i.e., when all the cell sizes n_{ij} , $i = 1, 2, \dots, a; j = 1, 2, \dots, b$ are the same, the size-adapted-weight method reduces to the equal-weight method.

It is now clear that each of the testing problems associated with the three null hypotheses (8) can be equivalently expressed in the form of the general linear hypothesis testing (GLHT) problem (10) as described in the next subsection with $\mathbf{C}$ respectively being

$$\mathbf{C}_a = \mathbf{H}_a \mathbf{A}_a, \quad \mathbf{C}_b = \mathbf{H}_b \mathbf{A}_b, \quad \mathbf{C}_{ab} = \mathbf{H}_{ab} \mathbf{A}_{ab}. \quad (9)$$

2.2 Modified Bartlett Test for Linear Hypotheses

Using the cell mean vector $\boldsymbol{\mu}$ defined in the previous subsection, we can write the GLHT problem under the heteroscedastic two-way ANOVA model (2) as:

$$H_0 : \mathbf{C} \boldsymbol{\mu} = \mathbf{c}, \quad \text{vs} \quad H_1 : \mathbf{C} \boldsymbol{\mu} \neq \mathbf{c}, \quad (10)$$

where $\mathbf{C} : q \times (ab)$ is a known coefficient matrix of full rank with $\text{rank}(\mathbf{C}) = q$ and $\mathbf{c} : q \times 1$ is a known constant vector, often specified as $\mathbf{0}$. The above GLHT problem is very general. In fact, the contrast matrices $\mathbf{H}_a$, $\mathbf{H}_b$, and $\mathbf{H}_{ab}$ in $\mathbf{C}_a$, $\mathbf{C}_b$, and $\mathbf{C}_{ab}$ can be replaced by any other proper contrast or non-contrast matrices. For example, to test $H_0 : \alpha_1 - .5\alpha_3 - .5\alpha_7 = 0$, we only need set $\mathbf{c} = 0$ and $\mathbf{C} = [\mathbf{e}_{1,a} - .5\mathbf{e}_{3,a} - .5\mathbf{e}_{7,a}] \mathbf{A}_a$, and to test $H_0 : \beta_3 = 4\beta_8$, we only need set $\mathbf{c} = 0$ and $\mathbf{C} = [\mathbf{e}_{3,b} - 4\mathbf{e}_{8,b}] \mathbf{A}_b$ where $\mathbf{A}_a$ and $\mathbf{A}_b$ are as defined in the previous subsection. To construct the test statistic for (10), we denote the usual unbiased estimators of the cell means and cell variances of the random sample (1) as:

$$\hat{\mu}_{ij} = n_{ij}^{-1} \sum_{k=1}^{n_{ij}} y_{ijk}, \quad \hat{\sigma}_{ij}^2 = (n_{ij} - 1)^{-1} \sum_{k=1}^{n_{ij}} (y_{ijk} - \hat{\mu}_{ij})^2, \quad i = 1, \dots, a; j = 1, \dots, b. \quad (11)$$

Set $\hat{\boldsymbol{\mu}} = [\hat{\mu}_{11}, \dots, \hat{\mu}_{1b}, \dots, \hat{\mu}_{a1}, \dots, \hat{\mu}_{ab}]^T$ as the estimator of $\boldsymbol{\mu}$. Then $\hat{\boldsymbol{\mu}} \sim N_{ab}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = \text{diag}(\frac{\sigma_{11}^2}{n_{11}}, \frac{\sigma_{12}^2}{n_{12}}, \dots, \frac{\sigma_{1b}^2}{n_{1b}}, \dots, \frac{\sigma_{a1}^2}{n_{a1}}, \frac{\sigma_{a2}^2}{n_{a2}}, \dots, \frac{\sigma_{ab}^2}{n_{ab}})$. Since $\mathbf{C} \hat{\boldsymbol{\mu}} - \mathbf{c} \sim N_q(\mathbf{C} \boldsymbol{\mu} - \mathbf{c}, \mathbf{C} \boldsymbol{\Sigma} \mathbf{C}^T)$, we can construct the following Wald-type statistic:

$$T = (\mathbf{C} \hat{\boldsymbol{\mu}} - \mathbf{c})^T (\mathbf{C} \hat{\boldsymbol{\Sigma}} \mathbf{C}^T)^{-1} (\mathbf{C} \hat{\boldsymbol{\mu}} - \mathbf{c}), \quad (12)$$

where $\hat{\boldsymbol{\Sigma}} = \text{diag}(\frac{\hat{\sigma}_{11}^2}{n_{11}}, \frac{\hat{\sigma}_{12}^2}{n_{12}}, \dots, \frac{\hat{\sigma}_{1b}^2}{n_{1b}}, \dots, \frac{\hat{\sigma}_{a1}^2}{n_{a1}}, \frac{\hat{\sigma}_{a2}^2}{n_{a2}}, \dots, \frac{\hat{\sigma}_{ab}^2}{n_{ab}})$. To construct the MB test, following Yanagihara and Yuan (2005), we re-express T as:

$$T = \mathbf{z}^T \mathbf{W}^{-1} \mathbf{z}, \quad (13)$$

where $\mathbf{z} = (\mathbf{C} \boldsymbol{\Sigma} \mathbf{C}^T)^{-1/2} (\mathbf{C} \hat{\boldsymbol{\mu}} - \mathbf{c})$, $\mathbf{W} = \mathbf{H} \hat{\boldsymbol{\Sigma}} \mathbf{H}$, and $\mathbf{H} = (\mathbf{C} \boldsymbol{\Sigma} \mathbf{C}^T)^{-1/2} \mathbf{C}$. It is easy to see that under the null hypothesis, $\mathbf{z} \sim N_q(\mathbf{0}, \mathbf{I}_q)$. Let $n_{\min} = \min_{i=1}^a \min_{j=1}^b n_{ij}$ and $n_{\max} = \max_{i=1}^a \max_{j=1}^b n_{ij}$ denote the minimum and maximum cell sizes respectively. To study the asymptotic null distribution of T , we impose the following condition:

$$\text{As } n_{\min} \rightarrow \infty, \quad \frac{n_{ij}}{n_{\min}} \rightarrow r_{ij} < \infty, \quad i = 1, 2, \dots, a; j = 1, 2, \dots, b. \quad (14)$$

This condition requires that all the cell sizes tend to infinity at about the same rate so that $n_{\min}(\mathbf{C}\Sigma\mathbf{C}^T)$ tends to a non-singular matrix as $n_{\min} \rightarrow \infty$. We then have the following result.

Theorem 1 Under the condition (14) and H_0 , as $n_{\min} \rightarrow \infty$, we have $T \xrightarrow{L} \chi_q^2$.

The above theorem shows that T asymptotically follows a χ^2 -distribution with q degrees of freedom. Similar results for some special $\mathbf{C}$ must have appeared in the literature but we here show that it is true for the GLHT problem (10). Based on Theorem 1, we can construct a χ^2 -test. However, it is well known that a χ^2 -test for (10) usually converges very slowly. In fact, from the proof of Theorem 1 in the Appendix, we can see that the convergence rate of T to χ_q^2 is of $O_p(n_{\min}^{-1/2})$. This indicates that the χ^2 -test based on T can be very inaccurate when $n_{\min}$ is small or moderate. To improve the convergence rate of T , the well-known Bartlett correction may be used. Recently, Fujikoshi (2000) proposed a modified Bartlett (MB) correction which can be used to improve the χ^2 -test further. We feel that this MB correction can be generally used to improve an asymptotical χ^2 -test provided that the asymptotical expressions for the mean and variance of the associated test statistic are available and follow some required structures. In this paper, we aim to illustrate how this technique can be used for (10). Let $\mathbf{h}_{ij} = (\mathbf{C}\Sigma\mathbf{C}^T)^{-1/2}\mathbf{c}_{ij}$, $i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$ where $\mathbf{c}_{ij}$ is the (ij) -th column of $\mathbf{C}$. To apply the MB correction to T , we need the following result.

Theorem 2 Under the condition (14) and H_0 , as $n_{\min} \rightarrow \infty$, we have

$$E(T) = q(1 + \frac{\alpha_1}{n_{\min}}) + O(n_{\min}^{-2}) \text{ and } E(T^2) = q(q+2)(1 + \frac{\alpha_2}{n_{\min}}) + O(n_{\min}^{-2}), \quad (15)$$

where $\alpha_1 = \frac{2n_{\min}}{q}\Delta$, $\alpha_2 = \frac{(14+4q)n_{\min}}{q(q+2)}\Delta$, and $\Delta = \sum_{i=1}^a \sum_{j=1}^b \left[\frac{\sigma_{ij}^2}{n_{ij}} \mathbf{h}_{ij}^T \mathbf{h}_{ij} \right]^2 / (n_{ij} - 1)$. In addition, we have

$$\frac{q^2}{(n_{\max} - 1)ab} \leq \Delta \leq \frac{q}{n_{\min} - 1}. \quad (16)$$

Theorem 2 allows us to apply the MB correction to T using the log-transformation $T_{\text{MB}} = (n_{\min}\beta_1 + \beta_2) \log(1 + \frac{T}{n_{\min}\beta_1})$, where $\beta_1 = \frac{2}{\alpha_2 - 2\alpha_1}$ and $\beta_2 = \frac{(q+2)\alpha_2 - 2(q+4)\alpha_1}{2(\alpha_2 - 2\alpha_1)}$. By Fujikoshi (2000), we have $T_{\text{MB}} \xrightarrow{L} \chi_q^2$ as $n_{\min} \rightarrow \infty$ with $E(T_{\text{MB}}) = q + O(n_{\min}^{-2})$ and $E(T_{\text{MB}}^2) = q(q+2) + O(n_{\min}^{-2})$. On contrast, by Theorem 2, we only have $E(T) = q + O(n_{\min}^{-1})$ and $E(T^2) = q(q+2) + O(n_{\min}^{-1})$. That is, the first two moments of T_{MB} converge to the first two moments of χ_q^2 much faster than those of T do. It is then expected that T_{MB} converges to χ_q^2 much faster than T does (Fujikoshi 2000).

In real data analysis, β_1 and β_2 are unknown. They must be estimated from the data. Good estimators can be obtained as long as a good estimator of Δ is supplied. By Theorem 2, a natural estimator for Δ is:

$$\hat{\Delta} = \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\sigma}_{ij}^2}{n_{ij}} \hat{\mathbf{h}}_{ij}^T \hat{\mathbf{h}}_{ij} \right]^2 = \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\sigma}_{ij}^2}{n_{ij}} \mathbf{c}_{ij}^T (\mathbf{C}\hat{\Sigma}\mathbf{C}^T)^{-1} \mathbf{c}_{ij} \right]^2. \quad (17)$$

The estimators $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1$, and $\hat{\beta}_2$ are then obtained accordingly. By the law of large numbers, it is classical to show that $\hat{\sigma}_{ij} \rightarrow \sigma_{ij}^2$, $i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$ almost surely as $n_{\min} \rightarrow \infty$. It follows that as $n_{\min} \rightarrow \infty$, we have $\frac{\hat{\Delta}}{\Delta}$, $\frac{\hat{\alpha}_i}{\alpha_i}$, and $\frac{\hat{\beta}_i}{\beta_i}$, $i = 1, 2$ all tend to 1. That is, they are ratio-consistent. Therefore, we continuously have

$$\hat{T}_{\text{MB}} = (n_{\min}\hat{\beta}_1 + \hat{\beta}_2) \log(1 + \frac{T}{n_{\min}\hat{\beta}_1}) \xrightarrow{L} \chi_q^2, \text{ as } n_{\min} \rightarrow \infty. \quad (18)$$

Some simple algebra gives that $n_{\min}\hat{\beta}_1 = \frac{q(q+2)}{3\hat{\Delta}}$ and $n_{\min}\hat{\beta}_1 + \hat{\beta}_2 = \frac{(q+2)(2q-\hat{\Delta})}{6\hat{\Delta}}$. From the proof of Theorem 2, one can easily see that the range of Δ given in (16) is also the range of $\hat{\Delta}$. Thus, provided $n_{\min} \geq 2$, we always have $n_{\min}\hat{\beta}_1 > 0$ and $n_{\min}\hat{\beta}_1 + \hat{\beta}_2 > 0$. This guarantees that $\hat{T}_{\text{MB}}$ is a nonnegative and monotonically increasing function of T .

The critical value of $\hat{T}_{\text{MB}}$ can be specified as $\chi_q^2(1 - \alpha)$ for any given significance level α . We reject the null hypothesis of (10) when this critical value is exceeded by $\hat{T}_{\text{MB}}$. The MB test can also be conducted via computing the P-value using the χ_q^2 -distribution (18) easily.

2.3 Some Desirable Properties of the MB Test

Notice that for each of the null hypotheses (8), the associated contrast matrix $\mathbf{H}_a, \mathbf{H}_b$ or $\mathbf{H}_{ab}$ is not unique. For example, $\tilde{\mathbf{H}}_a = \begin{pmatrix} -\mathbf{1}_{a-1}, \mathbf{I}_{a-1} \end{pmatrix}$ is also a contrast matrix for the first null hypothesis in (8). It is known from Kshirsagar (1972, Ch. 5, Sec. 4) that for any two contrast matrices $\tilde{\mathbf{H}}_*$ and $\mathbf{H}_*$ specifying the same null hypothesis, there is a nonsingular matrix $\mathbf{P}$ such that

$\tilde{H}_* = PH_*$, where the subscript $*$ denotes one of the subscripts a, b and ab , used for the three null hypotheses in (8). By (9), the C-matrix associated with the contrast matrix H_* can be expressed as $C_* = (H_*A_*)$. Let $\tilde{C}_*$ be the C-matrix associated with the contrast matrix $\tilde{H}_*$. Then we have $\tilde{C}_* = (\tilde{H}_*A_*) = (PH_*A_*) = PC_*$. Theorem 3 shows that the MB test is invariant to different choices of the contrast matrix for the same hypothesis.

Theorem 3 *The MB test is invariant when C and c in (10) are replaced with*

$$\tilde{C} = PC \text{ and } \tilde{c} = Pc, \quad (19)$$

respectively where P is any nonsingular matrix.

In practice, the observed cell responses are often re-centered or rescaled before any inference is conducted. Recentering and rescaling are special cases of the following affine transformation:

$$\tilde{y}_{ijk} = \lambda y_{ijk} + \xi, \quad k = 1, 2, \dots, n_{ij}; \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b, \quad (20)$$

where $\lambda \neq 0$ and ξ are two known constants.

Theorem 4 *The MB test is invariant under the affine transformation (20).*

Finally, we have the following result.

Theorem 5 *The MB test is invariant under different labelling schemes of the cell means $\mu_{ij}, i = 1, 2, \dots, a; j = 1, 2, \dots, b$.*

2.4 Parametric Bootstrap Test for Linear Hypotheses

The null distribution of the Wald-type statistic T (13) can also be approximated by a parametric bootstrap (PB) method. It is well known that under the heteroscedastic two-way ANOVA model (1), we always have $r_{ij} = \frac{\hat{\sigma}_{ij}^2}{\sigma_{ij}^2} \sim \frac{\chi_{n_{ij}-1}^2}{n_{ij}-1}, i = 1, 2, \dots, a; j = 1, 2, \dots, b$. Let $R = \text{diag}(r_{11}, \dots, r_{1b}, \dots, r_{a1}, \dots, r_{ab})$. Then the distribution of R is known so that when Σ is known, the distribution of $W = H\hat{\Sigma}H^T \stackrel{d}{=} H\Sigma^{1/2}R\Sigma^{1/2}H$ is known where $H = (C\Sigma C^T)^{-1/2}C$ as defined previously and $X \stackrel{d}{=} Y$ means X and Y have the same distribution. The PB test works via approximating the null distribution of T by the distribution of the bootstrap pivotal statistic $T_b = z^T W_b^{-1} z$ where $z \sim N_q(0, I_q)$ and $W_b = \hat{H}\hat{\Sigma}^{1/2}R\hat{\Sigma}^{1/2}\hat{H}^T$ with $\hat{H} = (C\hat{\Sigma}C^T)^{-1/2}C$. That is, it rejects the null hypothesis when $P(T_b > T_{obs}) < \alpha$ where α is the given significance level and T_{obs} is the observed value of T as given in (12). The left-hand side probability has to be evaluated using Monte Carlo via simulating (z, R) a large number of times. It is worthwhile to mention that we can show that Theorems 3, 4, and 5 also hold for the PB test.

3 Simulation Studies

In this section, we aim to compare the classical F, PB, and MB tests for two-way ANOVA via simulations. We implemented the three tests using the Matlab software. Upon an email request, readers can get the related Matlab codes from the corresponding author.

We first describe how the data in a two-way ANOVA model are generated. We do not need to directly generate the raw data; rather we need only to generate the sample cell means and variances using the given cell sizes, population means and variances since all the three tests are conducted using the sample cell sizes, means and variances only. For a given cell size vector $n = [n_{11}, n_{12}, \dots, n_{ab}]$, a cell mean vector $\mu = [\mu_{11}, \mu_{12}, \dots, \mu_{ab}]$ and a cell variance vector $\sigma^2 = [\sigma_{11}^2, \sigma_{12}^2, \dots, \sigma_{ab}^2]$, a sample cell mean vector $\hat{\mu} = [\hat{\mu}_{11}, \hat{\mu}_{12}, \dots, \hat{\mu}_{ab}]$ and a sample cell variance vector $\hat{\sigma}^2 = [\hat{\sigma}_{11}^2, \hat{\sigma}_{12}^2, \dots, \hat{\sigma}_{ab}^2]$ can be generated as $\hat{\mu}_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2/n_{ij})$ and $\hat{\sigma}_{ij}^2 \sim \frac{\sigma_{ij}^2}{n_{ij}-1} \chi_{n_{ij}-1}^2, i = 1, 2, \dots, a; j = 1, 2, \dots, b$. An illustrative example with $a = b = 4$ is presented in Table 1 in which the cell sizes, population means and standard deviations are listed in the columns associated with n_{ij}, μ_{ij} , and σ_{ij} respectively while the generated cell means and standard deviations are listed in the columns associated with $\hat{\mu}_{ij}$ and $\hat{\sigma}_{ij}$ respectively.

As an example, we applied the F, PB, and MB tests to the data in Table 1 in checking if there is an interaction effect between the two factors. In all the computations, the size-adapted-weight method was used. The P-values of the PB and MB tests were obtained as 55.5% and 56.1% respectively, suggesting that there is no strong evidence to reject the null hypothesis. However, the P-value for the classical F-test is 3.98%, suggesting a strong rejection for the null hypothesis. Therefore, in this example, the F-test produced a misleading conclusion while the PB and MB tests did not. This is because according to Table 1, the true cell means are all 0 so that the true interaction effect between the two factors is actually 0. This example shows the danger to apply the F-test blindly in a heteroscedastic two-way ANOVA model. In fact, this is not a single such case. The last row of Table 2 shows that at the nominal

TABLE 1 AN ILLUSTRATIVE EXAMPLE FOR DATA GENERATION IN A TWO-WAY ANOVA MODEL

$[i, j]$	n_{ij}	μ_{ij}	σ_{ij}	$\hat{\mu}_{ij}$	$\hat{\sigma}_{ij}$	$[i, j]$	n_{ij}	μ_{ij}	σ_{ij}	$\hat{\mu}_{ij}$	$\hat{\sigma}_{ij}$
[1, 1]	4	0	6	0.404	3.200	[3, 1]	4	0	6	-0.367	6.693
[1, 2]	8	0	1	-0.095	0.855	[3, 2]	8	0	1	-0.168	1.028
[1, 3]	12	0	1	0.211	0.764	[3, 3]	12	0	1	0.093	0.814
[1, 4]	20	0	1	-0.074	1.187	[3, 4]	20	0	1	0.156	0.961
[2, 1]	4	0	6	-3.574	2.706	[4, 1]	4	0	6	-2.254	3.212
[2, 2]	8	0	1	-0.520	0.857	[4, 2]	8	0	1	0.255	0.514
[2, 3]	12	0	1	0.351	0.706	[4, 3]	12	0	1	0.140	0.685
[2, 4]	20	0	1	-0.102	0.719	[4, 4]	20	0	1	0.219	1.066

TABLE 2 EMPIRICAL SIZES AND POWERS OF THE F, PB AND MB TESTS FOR VARIOUS CELL SIZES AND PARAMETER CONFIGURATIONS

σ	n	$\delta = 0$			$\delta = 2$			$\delta = 4$			$\delta = 6$		
		F	PB	MB	F	PB	MB	F	PB	MB	F	PB	MB
(1, 1, 1, 1) ₄	(11, 11, 11, 11) ₄	.050	.049	.044	.222	.203	.193	.819	.774	.771	.994	.990	.994
	(10, 20, 20, 30) ₄	.050	.049	.051	.374	.342	.354	.973	.962	.967	1.00	1.00	1.00
	(4, 8, 12, 20) ₄	.053	.049	.049	.184	.145	.133	.700	.569	.562	.985	.937	.944
(1, 2, 3, 4) ₄	(11, 11, 11, 11) ₄	.078	.054	.048	-	.062	.073	-	.156	.149	-	.365	.342
	(10, 20, 20, 30) ₄	.019	.050	.044	-	.085	.100	-	.304	.308	-	.669	.673
	(4, 8, 12, 20) ₄	.005	.041	.049	-	.065	.062	-	.153	.162	-	.345	.309
(1, 1, 1, 4) ₄	(11, 11, 11, 11) ₄	.106	.037	.046	-	.094	.102	-	.375	.369	-	.774	.766
	(10, 20, 20, 30) ₄	.016	.048	.057	-	.142	.149	-	.535	.555	-	.931	.929
	(4, 8, 12, 20) ₄	.001	.044	.049	-	.084	.089	-	.226	.223	-	.516	.522
(1, 1, 1, 6) ₄	(11, 11, 11, 11) ₄	.130	.044	.047	-	.096	.097	-	.337	.347	-	.729	.735
	(10, 20, 20, 30) ₄	.015	.050	.054	-	.132	.137	-	.489	.500	-	.899	.904
	(4, 8, 12, 20) ₄	.003	.054	.053	-	.079	.078	-	.224	.218	-	.463	.478
(4, 3, 2, 1) ₄	(11, 11, 11, 11) ₄	.073	.046	.049	-	.072	.069	-	.152	.154	-	.360	.342
	(10, 20, 20, 30) ₄	.249	.052	.049	-	.084	.086	-	.238	.251	-	.561	.571
	(4, 8, 12, 20) ₄	.435	.037	.051	-	.056	.062	-	.086	.105	-	.164	.199
(4, 1, 1, 1) ₄	(11, 11, 11, 11) ₄	.116	.047	.044	-	.103	.104	-	.362	.370	-	.767	.771
	(10, 20, 20, 30) ₄	.423	.057	.049	-	.188	.193	-	.750	.751	-	.992	.993
	(4, 8, 12, 20) ₄	.566	.043	.054	-	.074	.086	-	.239	.280	-	.605	.677
(6, 1, 1, 1) ₄	(11, 11, 11, 11) ₄	.132	.050	.040	-	.110	.103	-	.344	.347	-	.731	.742
	(10, 20, 20, 30) ₄	.454	.049	.048	-	.178	.188	-	.737	.747	-	.987	.990
	(4, 8, 12, 20) ₄	.622	.040	.058	-	.068	.098	-	.221	.273	-	.578	.655

size $\alpha = 5\%$, in 3000 replications, using the same cell sizes, population cell means and standard deviations as listed in Table 1, the empirical sizes of the F, PB and MB tests are 62.2%, 4.0% and 5.8% respectively. That is, the times of the null hypothesis rejected by the F, PB and MB tests among 3000 runs are approximately 1866, 120, and 174 respectively.

To further evaluate the performance of the F, PB and MB tests, in Table 2, we list the empirical sizes and powers of the three tests for testing if the interaction effect of the two factors is significant for more cell sizes and parameter configurations. For this purpose, we set $\mu_{ij} = ij\delta/(ab)$, $i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$. This implies that only when $\delta = 0$, the null hypothesis is satisfied. The cell standard deviations and sizes are listed in the first two columns where we use w_r to denote the vector obtained via repeating w for r times, e.g., $(1, 3)_2 = (1, 3, 1, 3)$. The empirical sizes of the three tests are listed in the three columns associated with $\delta = 0$, and the empirical powers are listed in the columns associated with $\delta = 2, 4$ and 6 . All the empirical sizes and powers were computed based on 3000 replications and for the PB test, the bootstrapped P-values were computed using 1000 simulation runs. To get more accurate bootstrapped P-values, more simulation runs for the PB test were preferred but the required computational effort forced us not to implement it in this simulation study.

From Table 2, we first notice that both the PB and MB tests performed well and they were generally comparable in terms of size controlling and power. We then compare the PB and MB tests against the F test in more details. As expected, from Table 2, it is seen that for those cell variance homogeneity cases, the F test outperformed the PB and MB tests in terms of size controlling and power although the PB and MB tests did not lose too much power. This is not a surprise since the F test takes the cell variance homogeneity assumption into account while the PB and MB tests do not. However, for those cell variance heteroscedasticity cases, the PB and MB tests performed much better than the F test in terms of size controlling. The F-test is either too conservative (with empirical size as small as .1%) or too liberal (with empirical size as large as 62.2%). The F-test is not acceptable for heteroscedastic two-way ANOVA since this may frequently yield misleading conclusions in practice as in the example shown in Table 1. Since for these cell variance heteroscedasticity cases, the empirical sizes of the F test are very different from those of the PB and MB tests, it does not make too much sense to compare the powers of the F-test with those of the PB and MB tests. That is why the powers of the F-test are replaced with “-” in Table 2.

TABLE 3 P-VALUES ($\times 100$) OF THE F, PB AND MB TESTS FOR TESTING THE MAIN AND INTERACTION EFFECTS OF AGE AND SEX FOR THE BASSIN ANTICIPATORY TIMING DATA

data used up to Trial	age			sex			age $\times$ sex		
	F	PB	MB	F	PB	MB	F	PB	MB
1	85.1	84.3	84.3	82.6	80.3	80.5	25.7	18.3	17.9
2	54.7	50.9	51.2	99.9	99.9	99.9	8.21	3.04	2.98
3	51.6	48.2	48.7	72.5	69.0	69.8	8.14	3.59	3.40
4	29.0	26.7	26.3	58.6	55.2	54.8	1.84	0.52	0.50
5	16.9	14.9	14.8	32.1	27.7	27.4	0.72	0.13	0.15

From these two simulation studies, we conclude that the PB and MB tests are comparable in terms of size controlling and power but they generally outperform the F test in terms of size controlling. In view of the much computational effort required by the PB test and the difficulty for checking the equality of cell variances, the MB test should be considered as a preferred choice in two-way ANOVA regardless if the cell variances are equal or unequal.

4 An Example

In this section, for illustration and comparison, we applied the classical F, PB and MB tests to a real data set known as the bassin anticipatory timing data, collected for a study about the reaction times of human beings and their hand-eye coordinations. There are 113 subjects who visited the Human Systems Integration Laboratory, Monterey, California, on 21 October 1995. Each of the subjects participated in an experiment in which the subject anticipated the arrival of a light to a designated point and a light beam was broken. The difference between the arrival time of the light and the time the light beam was broken was recorded. The experiment was administered for 5 trials. For each subject, his/her age and sex were also recorded. For the details of the experiment and the data set, readers are referred to <http://www.statsci.org/data/general/bassin.html>. Figure 1 displays the scatter plot of the anticipatory times of the subjects against their ages. It is seen that the anticipatory times of the subjects with ages below 20 spread much more widely than those of the subjects with ages above 20. It is then natural to group the ages into two categories: "age below 20" and "age above 20" so that the bassin anticipatory timing data can be treated under a heteroscedastic two-way ANOVA model where the two factors are "age" and "sex", each with 2 levels.

Table 3 displays the P-values of the F, PB, and MB tests applied to test the main and interaction effects for age and sex. Five cases of

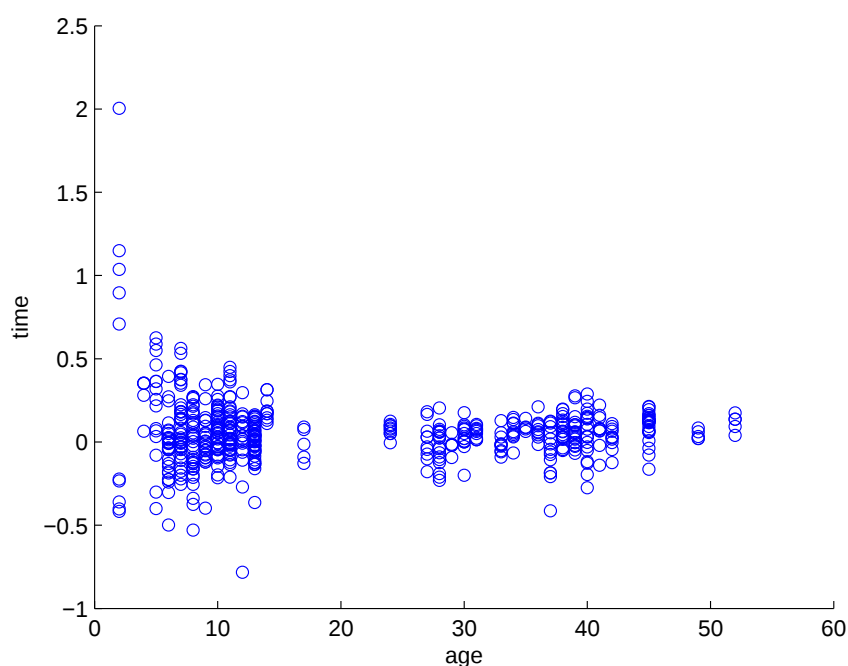


FIG. 1 THE SCATTER PLOT OF THE ANTICIPATORY TIMES OF THE SUBJECTS AGAINST THEIR AGES

the cell sizes were considered. In the k -th case for $k = 1, 2, 3, 4$, and 5 , only the data up to Trial k are involved. Different amount of the data involved allows us to compare the performance of the three tests at various cell sizes. For good accuracy, the P-values of the PB test were computed based on 10000 simulation runs. It is seen that for testing the main effects, the conclusions drawn from the three tests for five different sample sizes are consistent. It is then natural to compare the three tests for testing the interaction effect in more details. It is seen that for each of the cases considered, the P-value of the F test is much larger than those of the PB and MB tests, suggesting that the F-test are generally less powerful in this example at the presence of cell variance heteroscedasticity. It is seen that at 5% significance level, the PB and MB tests detected the significance of the interact effect between age and sex using the data only up to Trial 2 while the F test failed to do that until the data up to Trial 4 were involved. Therefore, from this example, we see that in real data analysis, the F-test has a larger risk in getting a misleading conclusion than the PB or MB tests do when the cell variance heteroscedasticity is present. Since the PB test is time-consuming, the MB test is generally preferred.

5 Conclusion

In this paper, we propose and study a MB test and a PB test for linear hypotheses in heteroscedastic two-way ANOVA in a unified manner. Simulation studies demonstrate that in terms of size controlling, both the tests outperform the classical F-test even for moderate heteroscedasticity and the classical F-test is either too conservative or too liberal when the homogeneity assumption is violated. Although both the tests are comparable in terms of size controlling, we recommend the MB test since it requires much less computational effort than the PB test does. Note that the weights used for the MB and MB tests can be specified by either the equal-weight method or the size-adapted-weight method. When the cell sizes are near the same, both the weight methods can be used. However, when the cell sizes are quite different, the size-adapted-weight method should be used so that the effect of the cell sizes can be taken into account.

ACKNOWLEDGMENTS

The work was supported by the NUS Academic Research Grant R-155-000-164-112. The authors thank two reviewers for their helpful comments and invaluable suggestions that help improve the paper substantially.

APPENDIX: Technical Proofs

Proof of Theorem 1 Under the given conditions, we have

$$\hat{\sigma}_{ij}^2 \stackrel{d}{=} \sigma_{ij}^2 \chi_{n_{ij}-1}^2 / (n_{ij} - 1), \quad i = 1, 2, \dots, a; j = 1, 2, \dots, b. \quad (\text{A.1})$$

It follows that $(\hat{\sigma}_{ij}^2 - \sigma_{ij}^2)/n_{ij} = O_p(n_{ij}^{-3/2})$, $i = 1, 2, \dots, a; j = 1, 2, \dots, b$. Thus $\hat{\Sigma} - \Sigma = O_p(n_{\min}^{-3/2})$. Noticing that $\Sigma = O(n_{\min}^{-1})$, we further have

$$\mathbf{R} = \mathbf{H}(\hat{\Sigma} - \Sigma)\mathbf{H}^T = O_p(n_{\min}^{-1/2}), \quad (\text{A.2})$$

where $\mathbf{H} = (\mathbf{C}\Sigma\mathbf{C}^T)^{-1/2}\mathbf{C}$ is defined in Section 2 and $\mathbf{H} = O(n_{\min}^{1/2})$. This implies that

$$\mathbf{W} = \mathbf{I}_q + \mathbf{H}(\hat{\Sigma} - \Sigma)\mathbf{H}^T = \mathbf{I}_q + \mathbf{R} = \mathbf{I}_q + O_p(n_{\min}^{-1/2}). \quad (\text{A.3})$$

Theorem 1 follows by Slutsky's theorem and noticing that under H_0 , $\mathbf{z}^T \mathbf{z} \sim \chi_q^2$.

Proof of Theorem 2 Notice that under H_0 , we have $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_q)$. Using the conditional expectation rule and some simple algebra leads to

$$\text{E}(T) = \text{E} \text{tr}(\mathbf{W}^{-1}) \quad \text{and} \quad \text{E}(T^2) = 2\text{E} \text{tr}(\mathbf{W}^{-2}) + \text{E} \text{tr}^2(\mathbf{W}^{-1}). \quad (\text{A.4})$$

From the proof of Theorem 1 and (A.2), we have that $\mathbf{W} = \mathbf{I}_q + \mathbf{R}$ with $\mathbf{R} = O_p(n_{\min}^{-1/2})$. Then we have $\mathbf{W}^{-1} = (\mathbf{I}_q + \mathbf{R})^{-1} = \mathbf{I}_q - \mathbf{R} + \mathbf{R}^2 - \mathbf{R}^3 + O_p(n_{\min}^{-2})$ and $\mathbf{W}^{-2} = (\mathbf{I}_q + \mathbf{R})^{-2} = \mathbf{I}_q - 2\mathbf{R} + 3\mathbf{R}^2 - 4\mathbf{R}^3 + O_p(n_{\min}^{-2})$. It is easy to see from (A.2) that $\text{E}(\mathbf{R}) = \mathbf{0}$ and $\text{Etr}(\mathbf{R}) = 0$. Thus

$$\begin{aligned} \text{Etr}(\mathbf{W}^{-1}) &= q + \text{Etr}(\mathbf{R}^2) - \text{Etr}(\mathbf{R}^3) + O(n_{\min}^{-2}), \\ \text{Etr}(\mathbf{W}^{-2}) &= q + 3\text{Etr}(\mathbf{R}^2) - 4\text{Etr}(\mathbf{R}^3) + O(n_{\min}^{-2}), \\ \text{Etr}^2(\mathbf{W}^{-1}) &= q^2 + \text{Etr}^2(\mathbf{R}) + 2q\text{Etr}(\mathbf{R}^2) - 2q\text{Etr}(\mathbf{R}^3) - 2\text{Etr}(\mathbf{R})\text{tr}(\mathbf{R}^2) + O(n_{\min}^{-2}). \end{aligned} \quad (\text{A.5})$$

To find $\text{Etr}(\mathbf{R}^2)$, $\text{Etr}^2(\mathbf{R})$, $\dots$, notice that by (A.2), we have

$$\mathbf{R} = \sum_{i=1}^a \sum_{j=1}^b \mathbf{h}_{ij} \mathbf{h}_{ij}^T u_{ij}, \quad (\text{A.6})$$

where $\mathbf{h}_{ij}$ is the (ij) -th column of $\mathbf{H}$ and $u_{ij} = \frac{\hat{\sigma}_{ij}^2 - \sigma_{ij}^2}{n_{ij}}$, for $i = 1, 2, \dots, a; j = 1, 2, \dots, b$. By (A.1), we have $E(u_{ij}) = 0$, $E(u_{ij}^2) = \frac{2\sigma_{ij}^4}{n_{ij}^2(n_{ij}-1)}$, and $E(u_{ij}^3) = \frac{8\sigma_{ij}^6}{n_{ij}^3(n_{ij}-1)^2}$, $i = 1, \dots, a; j = 1, \dots, b$, where we use $E(\chi_d^2/d - 1)^3 = 8/d^2$. Noticing that $u_{11}, u_{12}, \dots, u_{ab}$ are independent and by (A.6), we have

$$\begin{aligned} \text{Etr}(\mathbf{R}^2) &= \text{Etr}^2(\mathbf{R}) = \sum_{i=1}^a \sum_{j=1}^b (\mathbf{h}_{ij}^T \mathbf{h}_{ij})^2 E(u_{ij}^2) = 2\Delta, \\ \text{Etr}(\mathbf{R}^3) &= \text{Etr}(\mathbf{R}) \text{tr}(\mathbf{R}^2) = \sum_{i=1}^a \sum_{j=1}^b (\mathbf{h}_{ij}^T \mathbf{h}_{ij})^3 E(u_{ij}^3) \\ &= 8 \sum_{i=1}^a \sum_{j=1}^b \left[\frac{\sigma_{ij}^2}{n_{ij}} \mathbf{h}_{ij}^T \mathbf{h}_{ij} \right]^3 (n_{ij} - 1)^{-2} = O(n_{\min}^{-2}), \end{aligned} \quad (\text{A.7})$$

where $\Delta = \sum_{i=1}^a \sum_{j=1}^b \left[\frac{\sigma_{ij}^2}{n_{ij}} \mathbf{h}_{ij}^T \mathbf{h}_{ij} \right]^2 / (n_{ij} - 1)$. Combining (A.5) and (A.7) gives that $\text{Etr}(\mathbf{W}^{-1}) = q + 2\Delta + O(n_{\min}^{-2})$, $\text{Etr}(\mathbf{W}^{-2}) = q + 6\Delta + O(n_{\min}^{-2})$, and $\text{Etr}^2(\mathbf{W}^{-1}) = q^2 + (8q + 2)\Delta + O(n_{\min}^{-2})$. These, together with (A.4), yield that $E(T) = q \left[1 + \frac{\alpha_1}{n_{\min}} \right] + O(n_{\min}^{-2})$ and $E(T^2) = q(q + 2) \left[1 + \frac{\alpha_2}{n_{\min}} \right] + O(n_{\min}^{-2})$, where $\alpha_1 = \frac{2n_{\min}}{q} \Delta$ and $\alpha_2 = \frac{(14+4q)n_{\min}}{q(q+2)} \Delta$ as desired.

To find the lower and upper bounds of Δ as given in (16), for $i = 1, 2, \dots, a; j = 1, 2, \dots, b$, set $\mathbf{S}_{ij} = \frac{\sigma_{ij}^2}{n_{ij}} \mathbf{h}_{ij} \mathbf{h}_{ij}^T$. It is easy to see that $\mathbf{S}_{ij}$ is nonnegative and it has only one nonzero eigenvalue $\delta_{ij} = \frac{\sigma_{ij}^2}{n_{ij}} \mathbf{h}_{ij}^T \mathbf{h}_{ij}$. It is easy to verify that $\sum_{i=1}^a \sum_{j=1}^b \mathbf{S}_{ij} = \mathbf{I}_q$ so that $\sum_{i=1}^a \sum_{j=1}^b \delta_{ij} = \sum_{i=1}^a \sum_{j=1}^b \frac{\sigma_{ij}^2}{n_{ij}} \mathbf{h}_{ij}^T \mathbf{h}_{ij} = \text{tr}(\sum_{i=1}^a \sum_{j=1}^b \mathbf{S}_{ij}) = q$ and $\mathbf{I}_q - \mathbf{S}_{ij} = \sum_{(u,v) \neq (i,j)} \mathbf{S}_{uv}$. Therefore $\mathbf{I}_q - \mathbf{S}_{ij}$ is nonnegative, showing that the only nonzero eigenvalue δ_{ij} of $\mathbf{S}_{ij}$ is less than 1. It follows that $\Delta = \sum_{i=1}^a \sum_{j=1}^b \delta_{ij}^2 / (n_{ij} - 1) \leq q / (n_{\min} - 1)$ and $\Delta = \sum_{i=1}^a \sum_{j=1}^b \delta_{ij}^2 / (n_{ij} - 1) \geq \sum_{i=1}^a \sum_{j=1}^b \delta_{ij}^2 / (n_{\max} - 1) \geq q^2 / [(n_{\max} - 1)ab]$ where we have used the fact that for any nonnegative numbers $v_1, v_2, \dots, v_m$, we always have $\sum_{i=1}^m v_i^2 \geq (\sum_{i=1}^m v_i)^2 / m$. The theorem is then proved.

Proof of Theorem 3 From the definition (12) of T , it is easy to see that T is invariant under the transformation (19). Then by (18), we only need to show that $\hat{\Delta}$ is invariant under the transformation (19). For $i = 1, 2, \dots, a; j = 1, 2, \dots, b$, let $\mathbf{c}_{ij}$ and $\tilde{\mathbf{c}}_{ij}$ be the (ij) -th columns of the matrices $\mathbf{C}$ and $\tilde{\mathbf{C}} = \mathbf{P}\mathbf{C}$ respectively. Then it is easy to see that $\tilde{\mathbf{c}}_{ij} = \mathbf{P}\mathbf{c}_{ij}$, $i = 1, 2, \dots, a; j = 1, 2, \dots, b$. It follows from (17) that

$$\begin{aligned} \hat{\Delta} &= \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\sigma}_{ij}^2}{n_{ij}} \tilde{\mathbf{c}}_{ij}^T (\tilde{\mathbf{C}} \hat{\Sigma} \tilde{\mathbf{C}}^T)^{-1} \tilde{\mathbf{c}}_{ij} \right]^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\sigma}_{ij}^2}{n_{ij}} \mathbf{c}_{ij}^T \mathbf{P}^T (\mathbf{P} \mathbf{C} \hat{\Sigma} \mathbf{C}^T \mathbf{P}^T)^{-1} \mathbf{P} \mathbf{c}_{ij} \right]^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\sigma}_{ij}^2}{n_{ij}} \mathbf{c}_{ij}^T (\mathbf{C} \hat{\Sigma} \mathbf{C}^T)^{-1} \mathbf{c}_{ij} \right]^2 = \hat{\Delta}, \end{aligned}$$

as desired. The theorem is proved.

Proof of Theorem 4 This theorem is proved if we can show that both T and $\hat{\Delta}$ are invariant under the affine transformation (20). Let μ_{ij}, σ_{ij}^2 and $\tilde{\mu}_{ij}, \tilde{\sigma}_{ij}^2$ respectively denote the cell mean and variance of $y_{ijk}, k = 1, 2, \dots, n_{ij}$ before and after the affine transformation (20). Then we have $\tilde{\mu}_{ij} = \lambda\mu_{ij} + \xi$ and $\tilde{\sigma}_{ij}^2 = \lambda^2\sigma_{ij}^2$. It follows that $\mu_{ij} = \lambda^{-1}(\tilde{\mu}_{ij} - \xi)$. As we define the mean vector $\boldsymbol{\mu}$ and the variance matrix $\boldsymbol{\Sigma}$ in Section 2, we define $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\Sigma}}$ similarly. Then we have $\tilde{\boldsymbol{\mu}} = \lambda\boldsymbol{\mu} + \boldsymbol{\xi}$ and $\tilde{\boldsymbol{\Sigma}} = \lambda^2\boldsymbol{\Sigma}$, where $\boldsymbol{\xi} = \xi \mathbf{1}_k$. It follows that the GLHT problem (10) can be equivalently expressed as $\tilde{H}_0 : \tilde{\mathbf{C}}\tilde{\boldsymbol{\mu}} = \tilde{\mathbf{c}}, \text{ vs } \tilde{H}_1 : \tilde{\mathbf{C}}\tilde{\boldsymbol{\mu}} \neq \tilde{\mathbf{c}}$, where $\tilde{\mathbf{C}} = \lambda^{-1}\mathbf{C}$ and $\tilde{\mathbf{c}} = \lambda^{-1}\mathbf{C}\boldsymbol{\xi} + \mathbf{c}$.

Let $\hat{\mu}_{ij}, \hat{\sigma}_{ij}^2$ and $\hat{\tilde{\mu}}_{ij}, \hat{\tilde{\sigma}}_{ij}^2$ respectively denote the unbiased estimators of the cell mean and variance of $y_{ijk}, k = 1, 2, \dots, n_{ij}$ before and after the affine transformation (20). Then it is easy to see that $\hat{\tilde{\mu}}_{ij} = \lambda\hat{\mu}_{ij} + \xi$ and $\hat{\tilde{\sigma}}_{ij}^2 = \lambda^2\hat{\sigma}_{ij}^2$. It follows that $\hat{\tilde{\boldsymbol{\mu}}} = \lambda\hat{\boldsymbol{\mu}} + \boldsymbol{\xi}$ and $\hat{\tilde{\boldsymbol{\Sigma}}} = \lambda^2\hat{\boldsymbol{\Sigma}}$. Using the above, we have $\tilde{\mathbf{C}}\hat{\tilde{\boldsymbol{\mu}}} - \tilde{\mathbf{c}} = \lambda^{-1}\mathbf{C}(\lambda\hat{\boldsymbol{\mu}} + \boldsymbol{\xi}) - (\lambda^{-1}\mathbf{C}\boldsymbol{\xi} + \mathbf{c}) = \mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c}$ and $\tilde{\mathbf{C}}\hat{\tilde{\boldsymbol{\Sigma}}}\tilde{\mathbf{C}}^T = \lambda^{-1}\mathbf{C}(\lambda^2\hat{\boldsymbol{\Sigma}})(\lambda^{-1}\mathbf{C})^T = \mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T$. Thus, both $\mathbf{C}\hat{\boldsymbol{\mu}} - \mathbf{c}$ and $\mathbf{C}\hat{\boldsymbol{\Sigma}}\mathbf{C}^T$ are affine-invariant. So is T .

We now turn to show that $\hat{\Delta}$ is invariant under the affine transformation (20). Let $\hat{\tilde{\Delta}}$ be the affine-transformed $\hat{\Delta}$. Then by (17), we have

$$\begin{aligned} \hat{\tilde{\Delta}} &= \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\tilde{\sigma}}_{ij}^2}{n_{ij}} \tilde{\mathbf{c}}_{ij}^T (\tilde{\mathbf{C}} \hat{\tilde{\boldsymbol{\Sigma}}} \tilde{\mathbf{C}}^T)^{-1} \tilde{\mathbf{c}}_{ij} \right]^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\lambda^2 \frac{\hat{\sigma}_{ij}^2}{n_{ij}} \lambda^{-1} \mathbf{c}_{ij}^T (\mathbf{C} \hat{\boldsymbol{\Sigma}} \mathbf{C}^T)^{-1} \lambda^{-1} \mathbf{c}_{ij} \right]^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\sigma}_{ij}^2}{n_{ij}} \mathbf{c}_{ij}^T (\mathbf{C} \hat{\boldsymbol{\Sigma}} \mathbf{C}^T)^{-1} \mathbf{c}_{ij} \right]^2 = \hat{\Delta}, \end{aligned}$$

as desired. The theorem is then proved.

Proof of Theorem 5 The theorem is proved if we can show that both T and $\hat{\Delta}$ are invariant under different labelling schemes of the means μ_{ij} , $i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$. Let $i_1, i_2, \dots, i_a$ and $j_1, j_2, \dots, j_b$ be any permutations of $1, 2, \dots, a$ and $1, 2, \dots, b$ respectively. Then it is easy to see that $C\hat{\mu} = \sum_{i=1}^a \sum_{j=1}^b c_{ij} \hat{\mu}_{ij} = \sum_{u=1}^a \sum_{v=1}^b c_{i_u j_v} \hat{\mu}_{i_u j_v}$ and $C\hat{\Sigma}C^T = \sum_{i=1}^a \sum_{j=1}^b n_{ij}^{-1} \hat{\sigma}_{ij}^2 c_{ij} c_{ij}^T = \sum_{u=1}^a \sum_{v=1}^b n_{i_u j_v}^{-1} \hat{\sigma}_{i_u j_v}^2 c_{i_u j_v} c_{i_u j_v}^T$, showing that $C\hat{\mu}$ and $C\hat{\Sigma}C^T$ are invariant under different labelling schemes of the cell means and so is T .

The invariance of $\hat{\Delta}$ under different labelling schemes of the population means follows from (17) directly by noticing that

$$\hat{\Delta} = \sum_{i=1}^a \sum_{j=1}^b (n_{ij} - 1)^{-1} \left[\frac{\hat{\sigma}_{ij}^2}{n_{ij}} c_{ij}^T (C\hat{\Sigma}C^T)^{-1} c_{ij} \right]^2 = \sum_{u=1}^a \sum_{v=1}^b (n_{i_u j_v} - 1)^{-1} \left[\frac{\hat{\sigma}_{i_u j_v}^2}{n_{i_u j_v}} c_{i_u j_v}^T (C\hat{\Sigma}C^T)^{-1} c_{i_u j_v} \right]^2,$$

and that $C\hat{\Sigma}C^T$ is invariant under different labelling schemes of the cell means. This completes the proof of the proposition.

REFERENCES

- [1] Ananda, Malwane M. A., and Weerahandi, Samaradasa. "Two-way ANOVA with unequal cell frequencies and unequal variances." *Statistica Sinica* 7 (1997): 631-646.
- [2] Anderson, Marti, and Braak, Cajo Ter. "Permutation tests for multi-factorial analysis of variance." *Journal of Statistical Computation and Simulation* 73 (2003): 85-113.
- [3] Bao, Phan and Ananda, Malwane M. A.. "Performance of two-way ANOVA procedures when cell frequencies and variances are unequal." *Communications in Statistics - Simulation and Computation* 30 (2001): 805-829.
- [4] Fujikoshi, Yasunori. "Two-way ANOVA models with unbalanced data." *Discrete Math* 116 (1993): 315-334.
- [5] Fujikoshi, Yasunori. "Transformations with improved chi-squared approximations." *Journal of Multivariate Analysis* 72 (2000): 249-263.
- [6] Harrar, Solomon W., and Bathke, Arne C.. "A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations." *Annals of the Institute of Statistical Mathematics* 64 (2010): 135-165.
- [7] James, G. S.. "Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown." *Biometrika* 41 (1954): 19-43.
- [8] Konietzschke, Frank, Bathke, Arne C., Harrar, Solomon W., and Pauly, Markus. "Parametric and nonparametric bootstrap methods for general MANOVA." *Journal of Multivariate Analysis* 140 (2015): 291-301.
- [9] Krutchkoff, R. G.. "Two-way fixed effects analysis of variance when the error variances may be unequal." *Journal of Statistical Computation and Simulation* 32 (1989): 177-183.
- [10] Kshirsagar, A. M.. "Multivariate Analysis." New York: Marcel Dekker, 1972.
- [11] Mu, Weiyang, Xiong, Shifeng, and Xu, Xingzhong "Generalized confidence regions of fixed effects in the two-way ANOVA." *Journal of Systems Science and Complexity* 21 (2008): 276-282.
- [12] Pauly, Markus, Brunner, Edgar, and Konietzschke, Frank "Asymptotic permutation tests in general factorial designs." *Journal of the Royal Statistical Society: Series B* 77 (2015): 461-473.
- [13] Smaga, Lukasz. "Wald-type statistics using 2-inverses for hypothesis testing in general factorial designs." *Statistics & Probability Letters* 107 (2015): 215-220.
- [14] Weerahandi, Samaradasa. "Exact Statistical Methods for Data Analysis." New York: Springer-Verlag, 1995.
- [15] Wilcox, Rand R.. "Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA model." *Journal of Educational and Behavioral Statistics* 14 (1989): 269-278.
- [16] Xu, Li-Wen, Yang, Fang-Qin, Abula, Aji'erguli, and Qin, Shuang. "A parametric bootstrap approach for two-way ANOVA in presence of possible interactions with unequal variances." *Journal of Multivariate Analysis* 115 (2013): 172-180.
- [17] Yanagihara, Hirokazu, and Yuan, Ke-Hai. "Three approximate solutions to the multivariate Behrens-Fisher problem." *Communications in Statistics - Simulation and Computation* 34 (2005): 975-988.
- [18] Zhang, Jin-Ting. "Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices." *Technometrics* 53 (2011): 426-439.

- [19] Zhang, Jin-Ting. "An approximate degrees of freedom test for heteroscedastic two-way ANOVA." *Journal of Statistical Planning and Inference* 142 (2012a): 336-346.
- [20] Zhang, Jin-Ting. "An Approximate Hotelling's T-square test for heteroscedastic one-way MANOVA." *Open Journal of Statistics*. 2 (2012b): 1-11.
- [21] Zhang, Jin-Ting, and Xiao, Shengning. "A note on the modified two-way MANOVA tests." *Statistics and Probability Letters* 82 (2012): 519-527.